Introduction to Machine Learning

Numerical analysis - University of Luxembourg

Exercises

Linear regression

Exercise 1. *Linear Regression for Housing Prices*

In this exercise, you will build a simple machine-learning model to predict house prices based on given features. Here is a small dataset for training and testing your model:

Size (m^2)	Bedrooms	Bathrooms	Garage Size	Age (Years)	Distance to Center (km)	Renovation status	Price ($1000s)
50	1	1	0	30	10	0	398.5
60	2	1	1	20	8	1	468.5
70	2	2	1	50	15	0	474.0
80	2	2	2	10	5	2	593.0
90	3	2	1	25	7	1	621.5
100	3	3	2	40	12	0	647.3
110	3	2	2	5	4	2	751.3
120	4	3	2	15	6	1	798.0
130	4	3	2	45	13	0	791.9
140	4	3	3	8	3	2	911.0

Legend for Garage size:
`0 = no garage, 1 = 1 car, 2 = more`
Legend for Renovation Status:
`0 = Old, 1 = Medium, 2 = New`

- **Load the data**: define the dataset as a Python dictionary and convert it to a Pandas DataFrame.
- **Preprocess the data**: normalize the features to bring them to a similar scale.
- **Split the data**: divide the dataset into training and testing sets (80% train, 20% test).
- **Train a model**: use Linear Regression from scikit-learn to predict house prices based on the features.
- **Evaluate the model**: calculate the Mean Squared Error (MSE) on the test set.

Non Linear regression

Exercise 2. *Non-Linear Regression for Housing Prices*

In this exercise, you will predict housing prices using non-linear regression on the California Housing dataset available in scikit-learn.

Steps to follow:

1. Load the dataset using "data = fetch_california_housing()" and inspect it
2. Train and evaluate two models:
    - Polynomial Regression: A simple non-linear extension of linear regression.
    - Decision Tree Regression: A flexible tree-based model.
3. Compare both models using Mean Squared Error (MSE).